DOCUMENT RESUME

ED 317 578 TM 014 627

AUTHOR	Lesgold, Alan M.; And Others
TITLE	Cognitive and Instructional Theories of Impasses in Learning. Final Report.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	Office of Naval Research, Arlington, Va.
REPORT NO	UPITT/LRDC/ONR/IMP-1
PUB DATE	7 Feb 90
CONTRACT	N00014-86-K-0361
NOTE	45p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Cognitive Processes; College Students; Computer Uses in Education; Experiments; Higher Education; *Instruction; Learning Plateaus; Learning Problems; *Learning Theories
IDENTIFIERS	*Learning Impasses

ABSTRACT

This is the final report of a series of 19 experiments designed to study impasses in the learning of skills with a strong perceptual component. Several series of experiments were designed with the purpose of producing experimentally manipulable impasses or plateaus in the course of learning. Over 200 subjects (humans and animals) in learning studies identified targets in various complex computer-presented displays. Among the factors manipulated were: (1) complexity; (2) noise; (3) salience; (4) biasing instructions; and (5) the distribution of target features across boundaries of displays. Impasses were produced, but patterns of impasse phenomena were not reproduced reliably enough to support or disconfirm a theory of impasses in learning. It is suggested that the best available tools for studying impasses in learning are probably the tools of comparative expertise (expert-novice) research, rather than those of the learning study. Five figures supplement the text. (Author/SLD)

University of Pittsburgh

Learning Research and Development Center

Final Report

Cognitive and Instructional Theories of Impasses in Learning

Alan M. Lesgold, Gareth Gabrys, and Maria Magone
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

Phone: 412-624-7046
FAX: 412-624-9149
Internet: al@unix.cis.pitt.edu

This is a final report of research sponsored by the Cognitive Sciences Program, Office of Naval Research, under Contract No. N00014-86-K-0361.

Final Report

Cognitive and Instructional Theories of Impasses in Learning

Alan M. Lesgold, Gareth Gabrys, and Maria Magone
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

Phone: 412-624-7045
FAX: 412-624-9149
Internet: al@unix.cis.pitt.edu

This is a final report of research sponsored by the Cognitive Sciences Program, Office of Naval Research, under Contract No. N00014-86-K-0361.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION: Unclassified

1b RESTRICTIVE MARKINGS:

2a SECURITY CLASSIFICATION AUTHORITY:

2b DECLASSIFICATION / DOWNGRADING SCHEDULE:

3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.

4 PERFORMING ORGANIZATION REPORT NUMBER(S): UPITT/LRDC/ONR/IMP-1

5 MONITORING ORGANIZATION REPORT NUMBER(S):

6a NAME OF PERFORMING ORGANIZATION: Learning Research & Development Center, Univ. of Pittsburgh

6b OFFICE SYMBOL (If applicable):

6c ADDRESS (City, State, and ZIP Code): Pittsburgh, Pennsylvania 15260

7a NAME OF MONITORING ORGANIZATION: Cognitive Science Program, Office of Naval Research (Code 1142CS)

7b ADDRESS (City, State, and ZIP Code): Arlington, VA 22217-5000

8a NAME OF FUNDING / SPONSORING ORGANIZATION:

8b OFFICE SYMBOL (If applicable):

8c ADDRESS (City, State, and ZIP Code):

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-86-K-0361

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR04206	RR04206-01	

11 TITLE (Include Security Classification): Cognitive and Instructional Theories of Impasses in Learning

12 PERSONAL AUTHOR(S): Alan M. Lesgold, Gareth Gabrys, and Maria Magone

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 1986 TO 1989	1990, February, 7	36

16 SUPPLEMENTARY NOTATION:

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
05	09	

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Impasse, Learning impasses; behavioral, blocking, noticing; Spectrograms, inward/outward projecting masks; consonant discrimination, vowel transfer, feature noticing.

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This is the final report of a series of experiments designed to study impasses in the learning of skills with a strong perceptual component. Several series of experiments were designed with the purpose of producing experimentally manipulable impasses or plateaus in the course of learning. Subjects in learning studies identified targets in various complex computer-presented displays. Among the factors manipulated were complexity, noise, salience, biassing instructions, and the distribution of target features across boundaries of displays. Impasses were produced, but patterns of impasse phenomena were not reproduced reliably enough to support of disconfirm a theory of impasses in learning.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL: Dr. Susan Chipman

22b TELEPHONE (Include Area Code): 202/696-4318

22c OFFICE SYMBOL: ONR 1142CS

DD Form 1473, JUN 86 — Previous editions are obsolete — SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

Abstract

This is the final report of a series of experiments designed to study impasses in the learning of skills with a strong perceptual component. Several series of experiments were designed with the purpose of producing experimentally manipulable impasses or plateaus in the course of learning. Subjects in learning studies identified targets in various complex computer-presented displays. Among the factors manipulated were complexity, noise, salience, biassing instructions, and the distribution of target features across boundaries of displays. Impasses were produced, but patterns of impasse phenomena were not reproduced reliably enough to support or disconfirm a theory of impasses in learning.

Contents

The Concept of a Learning Impasse 3

Theoretical Views of Impasses 5

The Behavioral View 5

The Feature Sampling View 6

The Developmental View 9

Summary of Experimental Efforts 11

Lost Plane Experiments: September 1986 - December 1986 . . . 11

Lost Animal Experiments: November 1986 - October 1987 . . . 14

Easy Animals and Hard Animals Experiments 14

Extended Practice Animals and Nanimals Experiments 15

Reversed Animals and Within Animals Experiments 16

LCD Experiment: September 1987 17

Spectrogram Learning Experiments: November 1987 - June 1989 18

Vowel Scaling Experiment and Scale-Learn-Scale Experiment 21

Vowel Transfer Experiment 23

Real Word Learning Experiment 24

Instructional Model Experiment 25

Consonant Discrimination Learning Experiment I 27

Consonant Discrimination Learning Experiment II 28

Consonant Discrimination Learning Experiment (Selection Task) 29

Consonant Discrimination Learning Experiment III 29

Conclusions 31

Available Software and Data 33

References 34

The Concept of a Learning Impasse

This project was motivated by experiences in prior work on medical expertise and its acquisition (Lesgold, 1984a,b; Lesgold, Rubinson et al., 1988). We found that medical diagnostic performance showed certain aspects of nonmonotone change with practice, and this led us to wonder whether learning could be enhanced by finding ways to avoid apparent plateaus and setbacks. The concept of learning plateaus has had a checkered history in psychology (cf. Keller, 1958), but the discussions of plateaus were very superficial, simply asserting that they resulted from poor behavioral engineering and would not occur in any sensible instructional setting. We felt that modern science and technology created many circumstances in which plateaus might occur, and we wanted to gain some explanatory and experimental control over the phenomenon.

Our experience with impasses in learning came from studies of radiological expertise (Lesgold, 1984 a,b; Lesgold, Rubinson et al., 1988) and especially from learning studies that we conducted near the end of the radiology studies. The first phenomenon we noticed occurred in studies using an expert-novice type of comparative paradigm. We had no real novices. Rather, we compared radiologists with five or more years of post-residency experience with two groups of residents having either less than two years of residency experience or more than two years. In those studies, we found that the more advanced group of residents were less successful than either the junior resident group or the senior staff group. While the numbers of subjects were small, the effects were consistent. In several cases, junior residents in one study were accidentally used later as senior residents in a second study; on the same films, they reverted from correct diagnoses earlier in their careers to incorrect diagnoses later.

We also conducted a number of training studies in which we taught people over hundreds of trials to "diagnose" artificially generated displays that were similar to chest x-ray pictures and based on a more-or-less accurate anatomical model of the chest. In these unpublished studies, we varied the amount of conceptual knowledge about the chest that was provided to subjects, and we found that subjects taught an appropriate mental model for the chest and its connection with the displays took as long or longer in

initial learning and showed no greater transfer to displays based on variations in the chest "diseases" on which the original displays were based (e.g., collapsed left upper lung instead of collapsed right middle lung) than subjects who did not receive the conceptual training. Further, some display types showed no learning over long periods of training (i.e., no movement above chance performance).

After reading some of the literature on non-monotone aspects of development and some of the concept learning literature, it became apparent to us that certain aspects of modern life create opportunities to view the world in ways that are more subject to learning impasses than might be the case in a more "natural" world. Our view has been, in essence, that impasses occur only in cases where (a) the situation to be understood or recognized is extremely complex, (b) the structure of features apparent in the situation does not map very directly onto any model of the world that the learner might have, and (c) the learner has not yet acquired any direct organization of the microfeatures of the situation into higher-order features that might have such a direct mapping into his/her conceptual model repertoire.

One example of such a situation is passive sonar image interpretation. Passive sonar images are distributions showing energy levels of different sound frequencies over time. The "objects" in such displays do not map directly onto the objects of the ocean environment. Rather, they map onto summations of sound producing activities. Further, each sound producing activity is likely to produce several unique "objects" in a distribution of spectral energy over time, and individual components of such "objects" may be closer to components of other "objects" than to each other. Accordingly, the potentially meaningful units according to the Gestalt rules may not be meaningful at all. Such situations seem likely to be artificial—based on some man-made artifices—rather than naturally occurring. They are not entirely novel, but they are certainly more common with new technologies. Other situations of this sort include 12-lead electrocardiograms, well logs from oil exploration studies, and densely-packed printed circuit and VLSI layouts.

We hoped to bring the impasse phenomena produced by such situations under experimental control, and that was the purpose of this

project. We were not entirely successful. Indeed, we asked ONR not to consider the optional third year for our contract, because we feel that significant progress must await the development of entirely different experimental approaches than those we took. After performing 19 experiments, we still find ourselves unable to demonstrate and control impasse phenomena adequately to meet our standards of empirical science. In the sections that follow, we summarize theoretical viewpoints of possible relevance, our many empirical studies, and our final conclusions.

Theoretical Views of Impasses

There are several levels at which one can view learning impasses. Clearly, they can be seen at the cognitive level hinted at in the discussion above, either fully within a theoretical stance based on mental models or from a developmental point of view. However, they might also be seen from a behavioral point of view or from a perceptual learning point of view, and certain aspects of these non-cognitive viewpoints seem worthy of note.

The Behavioral View

The conditioning literature contains references to certain cases in which stimulus patterns either are not conditionable to responses or else take a long time to become conditioned. Two related phenomena that have been reported are *overshadowing* and *blocking* (cf. Mackintosh, 1975). Both refer to situations in which one stimulus which is correlated with another cannot be conditioned to a response. Overshadowing is a phenomenon originally reported by Pavlov, in which a more salient stimulus, when conditioned to a response, prevents the conditioning of a less salient but equally relevant (i.e., predictive) stimulus to that response. For example, if a weak thermal stimulus is presented shortly before food is supplied, a dog will learn to salivate in response to that stimulus. However, if the thermal stimulus is always accompanied by a loud noise, only the noise will be conditioned.

Blocking is a term introduced by Kamin (1969) in which conditioning one stimulus to a response prevents later conditioning of a second element

after both are presented together. For example, if light is used to signal a shock and then later light and noise together signal the coming shock, the noise alone will not come to elicit any shock-related response. This phenomenon is similar to one seen in some of our experiments on voice spectrogram recognition described below.

Mackintosh (1975) suggested that a stimulus will be conditioned to the extent that it signals a change from what could have been predicted without it. Further, he theorized, stimuli that have no marginal predictive power become less conditionable. To the extent that a stimulus's predictive power is, or appears to the subject to be, stochastic, a change in predictive power will take time to notice. Hence, if Mackintosh is correct, a stimulus without predictive power that becomes predictive will initially suffer a period of slow learning because of the compounding of the partial reinforcement effect and the initially lower learning rate due to historically being low in marginal predictive capability.

The Feature Sampling View

The behavioral data just reviewed may seem of minimal relevance to impasses in cognitive learning, but it does prompt us to notice several aspects of the impasse situations we have examined and to better understand how those situations deviate from experimental paradigms that have been employed in studying plateaus and impasses. Concept learning experiments tend to use relatively simple displays. The most common type of experiment uses displays in which there are a small number of dimensions varied, each involving a small number of display features, e.g., single vs. double borders, square vs. triangle, one vs. two central forms, red vs. blue, etc. A second type of display form that has been used in experimental work is the random deviation from a prototype. The so-called Attneave Figure is such a form. To define each prototype, a set of randomly plotted points is connected to create a polygon. Instances of the prototype are created by introducing small random perturbations of the exact locations of the vertex points. Three instances of the same prototype are shown in Figure 1 below.

Attneave figures and the simple displays of concept learning experiments can be contrasted with the much more complex displays that

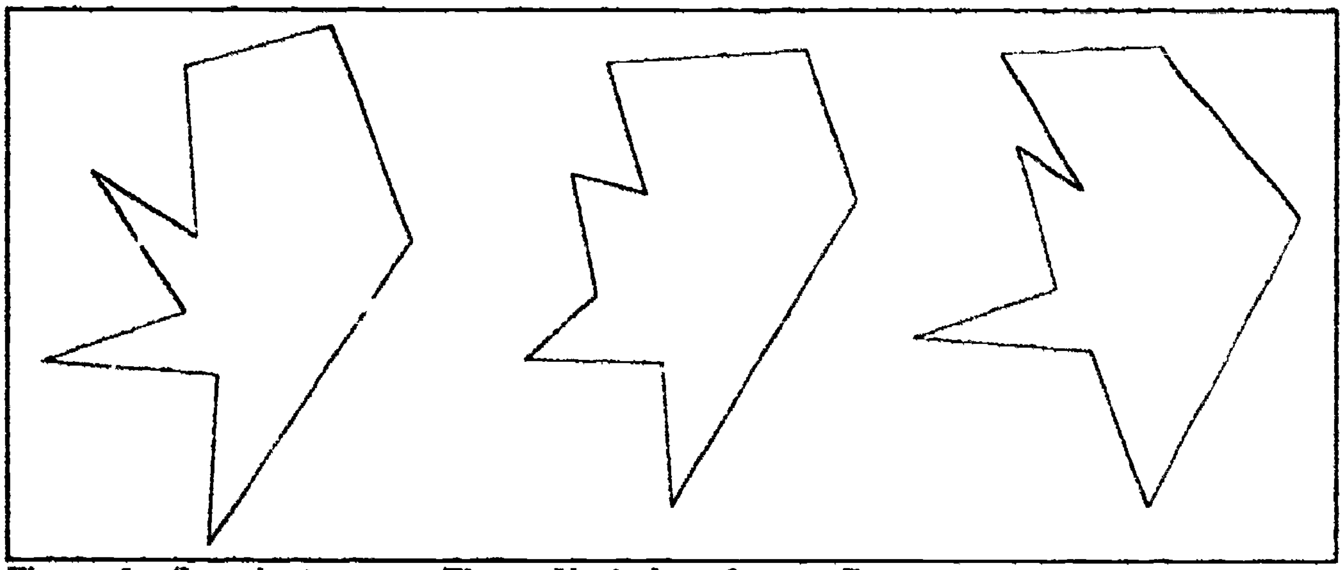

Figure 1. Sample Attneave Figure Variations from a Prototype.

were the target of this project, passive sonar displays, voice spectrograms, and the like. In the figures that have been used for experimental work, the features that might play a role in defining categories are relatively evident. In contrast, the meaningful features of the noisy artificial displays in which we were interested are very difficult to isolate. Sometimes, critical features or feature relationships are never noticed over the course of several hours of experimentation. In this respect, standard methodologies of concept learning, which look at the relative speed at which different kinds of concepts are acquired, and perceptual learning experiments which look at the relative speed at which different display types come to be recognized, were not suited to our goals. As will be seen below, when we used realistic stimuli, many subjects failed ever to learn what to notice. When we used simpler stimuli, we failed to get impasse effects.

The time needed to discover which features are relevant in a perceptual recognition learning task is an important measure. For example, Zeaman and House (1963; see also Fisher & Zeaman, 1973) found that retardates differed from normal subjects in how long it took them to notice relevant stimulus features. Once features were noticed by retardates, their improvement curves looked about the same as those for normal subjects. This motivates an experimental paradigm in which trials until learning starts to be evident is a basic measure. However, with the materials in which we were interested, such experiments proved impossible to run successfully. In order to be practical and yet of sufficient power, the experiments required

within-subject manipulations. However, when learning failed to occur at all for some cases, these within-subject studies were not entirely conclusive.

The difficulty problem makes it impossible to clearly separate two important potential causes of perceptual learning impasses. One is inability to notice critical features, as just discussed. A second, and one that we think is important (see the discussions below of our artificial voice spectrogram studies) is whether critical feature combinations consist of features that are all within the same meaningful region of a display or not. As a specific example, consider the case of voice spectrograms for syllables. In such displays, it is possible, and obviously meaningful, to parse the display into segments corresponding to individual phonemes. The display plots time on the x axis against frequency on the y axis, and it makes sense to split up the total time into the periods in which each of the phonemes of a syllable were uttered. However, since it also takes time for the speech apparatus to reconfigure from one phoneme to the next, some of the cues for identifying one phoneme are to be found in the features of the phoneme immediately before or after. For example, distinguishing /d/ from /g/ is generally difficult to impossible without examination of the features of the vowel that follows (as in *dig* vs. *gig*).

This is an example of the general problem, cited above, in which the apparent spatial components of a display do not map well onto the components of the events that gave rise to the display. Unfortunately, we failed to gain control over this kind of situation. While some of our final experiments demonstrate weakly that such a problem is significant, we could not control its emergence well enough to permit the kinds of instructional studies we wanted to carry out. This outcome is particularly discouraging because better theoretical apparatus is being developed for understanding how people come to discover the feature clusters that are relevant to a learning task. For example, Billman and Heit (1988) have simulated the effects of some very general, or weak, metacognitive methods of focused sampling of potential rules for mapping features and feature combinations onto categories, a significant step beyond the simple formulations of Zeaman (French & Zeaman, 1973; Zeaman & House, 1963).

The Developmental View

The developmental literature also provides quite a bit of theoretical power for dealing with learning impasses. Again, the problem is that we could not gain adequate experimental control to apply current theory. Stage theories of cognitive development are inherently theories of impasse, asserting that certain learning, possible at later stages of development, cannot occur earlier. In fact, the developmental literature is replete with examples of non-monotone learning curves, situations in which performance suffers setbacks, in terms of some fixed criterion, over the course of practice (Bowerman, 1982; Karmiloff-Smith, 1979; Karmiloff-Smith & Inhelder, 1974/1975; Klahr, 1982; Richards & Siegler, 1982; Stavy, Strauss, Orpaz, & Carmi, 1982; Strauss & Stavy, 1982). In fact, Strauss & Stavy (1982) listed five kinds of nonmonotone performance possibilities:

1. Movement from a practiced but inadequate mental representation of a task situation to a more powerful but less-well-practiced representation.

2. Uncoordinated combination of two different mental representation systems.

3. Using newly-learned rules that are correct for one situation in apparently related situations for which they are incorrect.

4. Having lower-order rules to deal with each of two task variables but not having the higher-order rules to coordinate these lower-order rules.

5. Having problems adapting a newly-acquired weak method to a specific situation for which a more domain-specific strong method must be evolved before the new metacognitive knowledge can be effective.

We believe that the problems faced by people trying to learn to recognize displays like passive sonar images and voice spectrograms do indeed involve mental representation inadequacies, but they are perhaps of a

slightly different character than has been examined in the developmental literature. The problem appears to be that in order to quickly apprehend these artificial displays, one must be able to recognize complex features that are not physically clustered according to the Gestalt laws (e.g., the features close together may not be related and ones far apart might be closely related). Generally, in order to handle such situations, one needs to be able to recognize the relevant lower-order features, to know parsing rules for sorting out which lower-order features cluster together, and to understand the meaning of the clusters.

This is not something that people are good at, in general. After all, the case of speech perception is remarkably similar. The superficial clustering, in terms of bursts of sound, for spoken language does not match word boundaries very well (e.g., *goo/d eve/ning* or *a/llon/s en/fants de/ la pa/tri/e*). Rather, we become highly practiced at matching these sound patterns to representations of the concepts to which they refer, even though that requires a highly specialized parsing. This parsing ability does not arise without extensive practice. Even moving from one language to another requires substantial practice. Further, in the speech understanding case, our own experience tells us that the study of vocabulary and grammar do not, themselves, permit understanding of the spoken word—one has to practice conversations extensively to learn to understand a new language as spoken. Prior reading knowledge certainly helps, but only to a point.

The time course of such practice makes it very difficult to conduct learning studies. As a result, much of developmental psychology involves comparisons of performance of different people selected from different points in the learning/development curve. Further, extensive interactions and verbal thinking-aloud protocols are often used. This is sufficient for characterizing the course of development, but it does not admit readily the possibility of studying systematically varied experience tracks. Small amounts of comparative ethnographic work have been done, but for the most part developmental methods are insufficient for studying the effects of various training interventions.

Nonetheless, we had hoped to use such methodologies as an adjunct to our experimental manipulations. Indeed, in some of the studies reported below, we did take protocols in order to better understand how subjects were

trying to learn to recognize various patterns. However, our failure to predictably generate impasse effects in experimentally tractable ways kept us from pursuing the developmental approach very far. We did, however, get some sense in a few of our studies of the ways in which subjects were trying to sort out what they were seeing and therefore of the mental models that they had for the domains we used.

Summary of Experimental Efforts

Since the fall of 1986, a total of 19 experiments were designed in which at least one subject was run. Because the experiments used displays generated by complex rules, all of the experiments were conducted on Xerox artificial intelligence workstations. The programs used to generate the displays and to conduct the experiments are available from the authors and will be sent without charge to anyone on the ONR Cognitive Science mailing list who requests them. The following is a summary of these experiments and their results. Individual reports of the experiments give more detailed descriptions of the experiments (see "Available Software and Data").

Our first attempts to produce reliable and experimentally tractable impasses used extremely noisy displays of known object form classes, such as animals and airplanes. We chose these displays in the hope that this would allow us to keep the tasks simple enough to fit standard experimental paradigms and time constraints. We then tried using displays that resembled the segmented digits used on LCD watches. Finally, we conducted an extensive series of studies using artificially created displays that resembled voice spectrograms.

Lost Plane Experiments: September 1986 - December 1986

Two experiments were conducted in which subjects studied three different drawings of military planes and then were given a series of visual search trials in which they were to identify the plane that appeared on the screen and its directional orientation (the latter a control for guessing). The planes were obscured by a moderate amount of random line noise (lines or curves of random length and orientation) and randomly strewn plane parts (wings or tails). The two versions of the experiment, called Easy Planes and

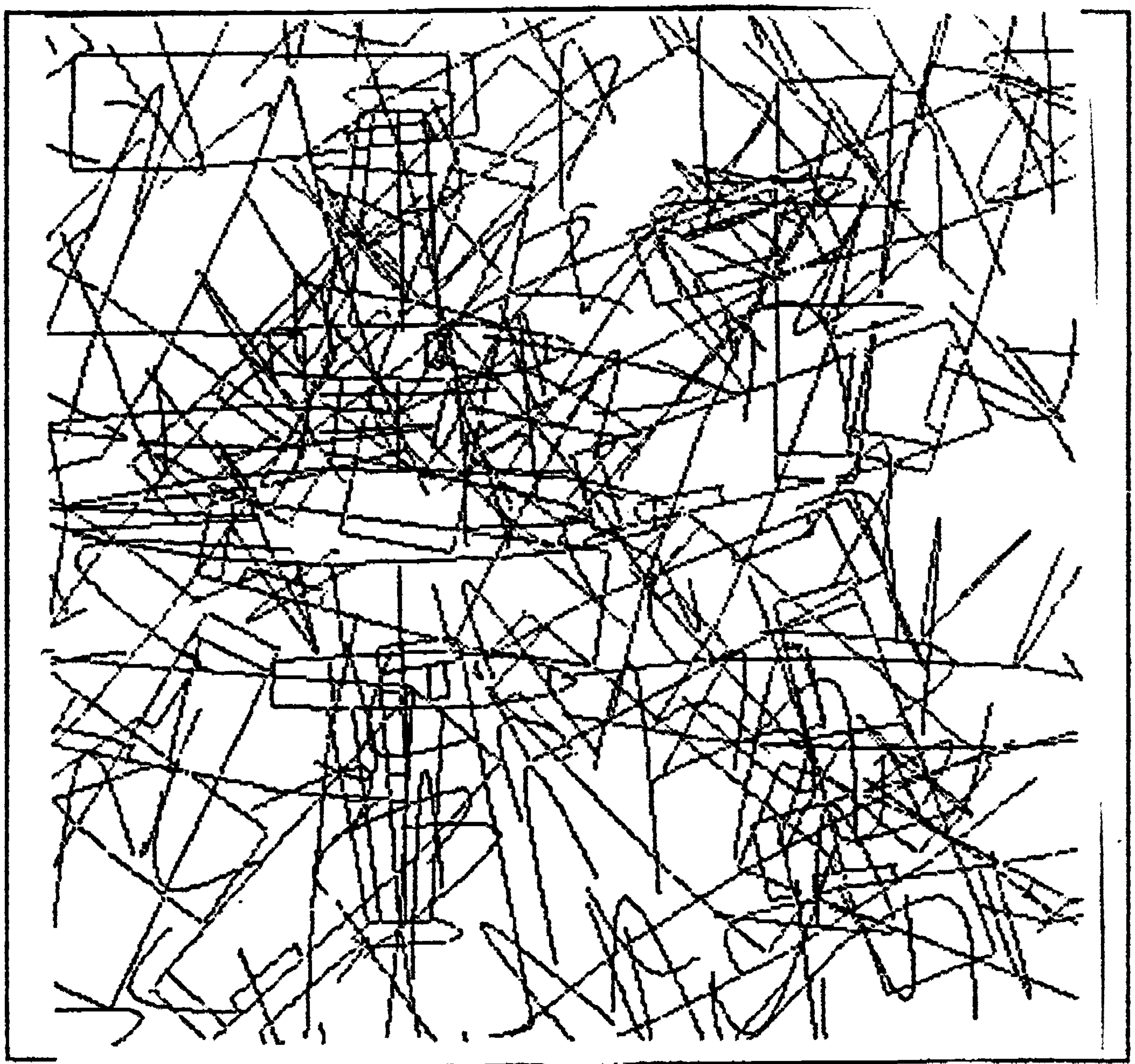

Figure 2. Easy plane facing east with wing noise.

Hard Planes, differed only in the amount of random line noise used. Figures 2 and 3 show examples of an easy and a hard case.

Method. There were three different plane silhouettes, and the task was to learn to identify which plane was hidden in the display. The manipulated variables for the experiments were the Plane Identity (A, B, or C), the Orientation of the plane (8 compass values), and the type of Plane Parts used as masking noise (either wings from Plane A, or tails from plane C). Combinations of these variables produced 48 different pictures which were presented to the subject in 4 blocks of 12 trials. Twenty subjects participated in the Easy Planes experiment, and six participated in the Hard Planes experiment.

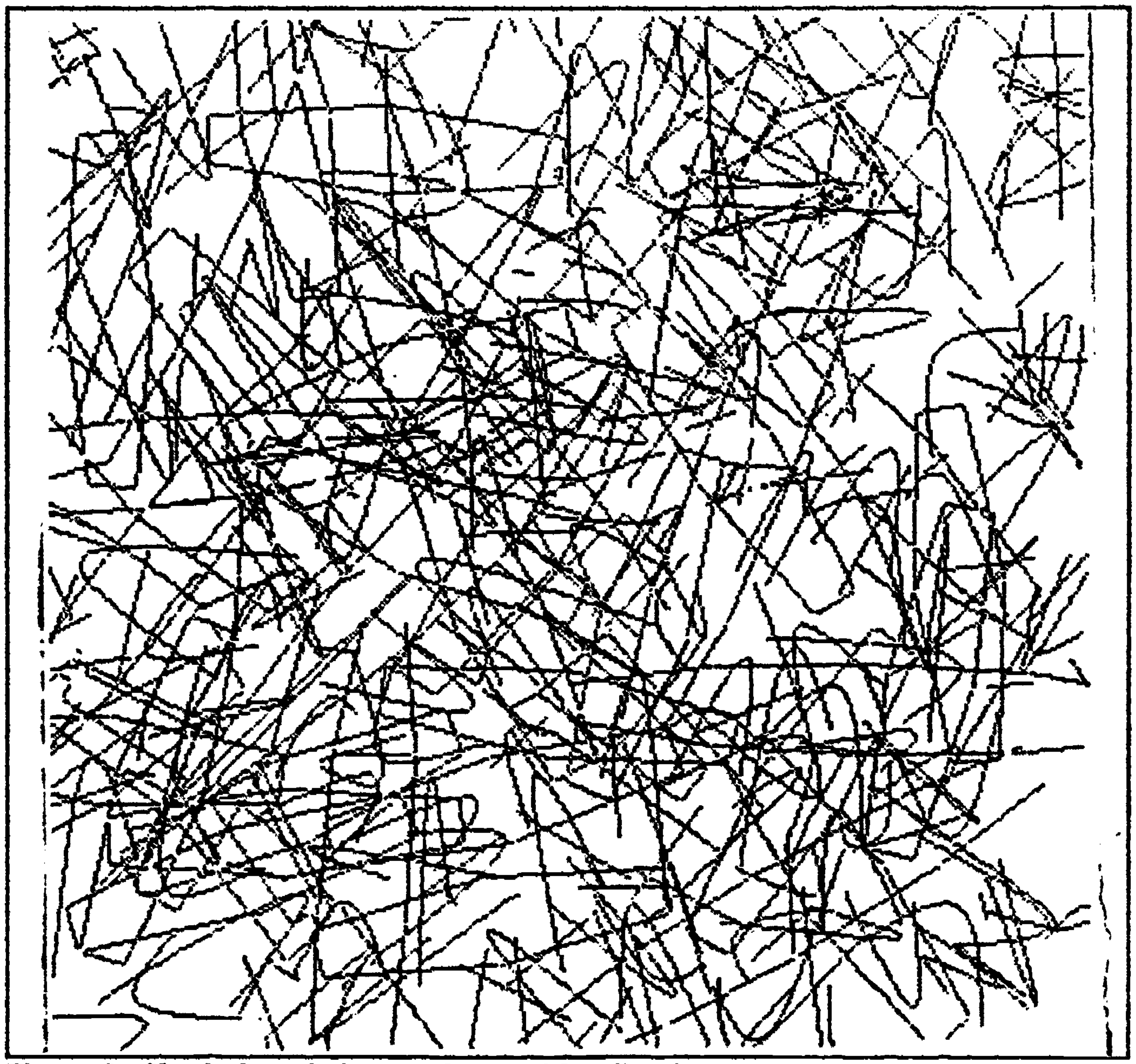

Figure 3. Hard plane facing northeast with tail noise.

Results. Because our focus was on reliably generating learning impasses,we could not fully control all variables. Specifically, the design of the experiments unsystematically confounded Orientation with Learning Block. Hence, a full factorial analysis could not be performed. This should be kept in mind when considering the following results. For the Easy Planes experiment, mean proportion correct over learning blocks increased linearly from 0.55 to 0.92 while response time decreased linearly from 33.82 seconds to 16.43 seconds. There were no systematic learning differences for the different Plane Identities or Parts Masks. For the Hard Planes experiment, mean proportion correct increased linearly from 0.44 to 0.79 over learning blocks as response time decreased from 55.27 to 41.45 seconds. Again no

systematic learning differences were observed for either Plane Identity or Parts Mask type. No learning impasses were observed.

Lost Animal Experiments: November 1986 - October 1987

The lost animals experiments were similar in principle to the lost planes experiments. Generally, subjects were shown outline drawings of five animals to study, and were then presented with several visual search trials where they were to identify an animal and specify its orientation. Altogether, seven lost animals experiments were conducted. These included manipulations of noise type (Easy Animals and Hard Animals), manipulation of the subject's advance knowledge of the animal shapes and identities (Free Response Animals), extended practice on the difficult animals task by the experimenters (Extended Animals, and Nanimals), and comparison of learning ability with parts masks which were inward projecting, where the parts could belong to animals within the picture, or outward projecting, where the parts could not belong to animals within the picture (Reversed Animals and Within Animals).

Easy Animals and Hard Animals Experiments

The Easy and Hard Animals experiments were basically the same in design as the Lost Plane experiments. Subjects viewed five outline drawings of animals and then performed a visual search task where they specified which animal was depicted and which orientation it faced. In the Easy Animals Experiment, the animals were shown with one of two types of random line noise: either straight lines or curved lines. In the Hard Animals experiment, the random line noise was augmented with a mask made up of animal parts (e.g., kangaroo tail, elephant trunk, etc.).

Method. The manipulated variables were Animal Identity (Penguin, Camel, Rhinoceros, Kangaroo, Elephant), Orientation (four primary compass values), and Noise Type (straight or curved). Combinations of these variables produced 40 different pictures which were shown to subjects in blocks of 10 trials. Sixteen subjects participated in each of the Easy and Hard Animals experiments, but no subject participated in both experiments.

Results. As was the case for the Lost Planes experiments, the Lost Animals experiments also unsystematically confounded Orientation with Learning Block. Hence, no full factorial analysis was possible. Keeping this in mind, the mean proportion correct for the Easy Animals experiment increased slightly with learning block. The values range from 0.80 to 0.89. At the same time, response time decreased from 15.76 seconds to 9.66 seconds with learning block. So, again there were no reliable impasse effects. No systematic learning differences between animals were found, but animals disguised in straight line noise were more often detected than animals disguised in curved noise. Straight line noise accuracy was at ceiling on all four learning blocks, but Curved line noise accuracy appeared to improve from 0.67 to 0.84.

The results for the Hard Animals experiment were that subjects performed only slightly above chance during the experiment and never improved (0.10 on block 1 to 0.11 on block 4; chance was 0.05). Subjects were only slightly more accurate on animals masked by straight line noise (0.13) than on animals masked by curved line noise (0.09). It was this finding of an apparent impasse that kept us persisting with the animal detection studies.

Extended Practice Animals and Nanimals Experiments

To discover whether the Hard Animals task could be learned, the experimenters performed the task over several sessions. In the Extended Practice experiment, two experimenters (MM and GG) familiar with the task performed it 8 times. In the Nanimals experiment, an experimenter (JT) unfamiliar with the task performed it 20 times. In this latter experiment, different parts masks were used on each trial to prevent improvement due to learning the position of the distractors.

Method. The experiment was the standard Hard Animals experiment described above. For the Nanimals experiment, the animal parts mask was changed on each problem to prevent the position of the distractors from being learned. However, the same set of masks were used on each session.

Results. Again, no factorial analysis of the results will be presented, but overall improvement in accuracy and response time was found. That is, given adequate practice, learning occurred continuously without impasse. For the Extended Practice experiment, one subject (GG) began with ceiling accuracy and decreased in response time from a mean of 27.72 seconds on the first block of the first session to a mean of 4.83 seconds on the final block of the 8th session. The other subject (MM) reached ceiling accuracy on the second session and decreased in response time from a mean of 67.42 seconds on the first block of the second session to 11.91 seconds on the last block of the 8th session.

For the Nanimals experiment, the subject (JT) achieved an accuracy of 0.10 on the first session (comparable to the performance of subjects in the Hard Animals experiment) and reached ceiling accuracy by about the 7th session. From this point, response time decreased from 26.34 seconds on the first block of the 7th session to 9.80 seconds on the final block of the 20th session. Again, the basic finding is that the task, too difficult for the time constraints of ordinary laboratory experimentation, showed no real impasses when adequate training time was given.

Reversed Animals and Within Animals Experiments

Even though continuous learning took place if enough trials were given, the hard animals tasks could, on the right time scale, be seen as involving impasses in learning, at least for the less-motivated subjects we recruited (relative to our own staff in the extended studies). So, we tried to find controlled means for making the difficulty of the hard animals conditions come and go. These experiments examined whether the search difficulty created by the animal parts mask (as was found in the Hard Animals experiment) was due to subjects being misled into examining the parts contained in the mask. The parts mask used by the Hard Animals experiment located animal parts so that if the rest of the animal were attached to the part, the whole animal would appear within the stimulus picture. For this reason, the mask was called "inward projecting." A second mask was designed which located the same parts so that if the rest of the animal were attached to the part, most of the animal would be located outside of the stimulus picture. This second mask was called "outward

projecting." The reasoning behind the experiments was that if subjects were testing part hypotheses during their search, they should be more disrupted by the inward projecting mask, whose parts they would have to test, than by the outward projecting mask, whose parts they should be able to quickly reject as potential targets. The two experiments differ in that the Reversed Animals experiment uses a between-subject design while the Within Animals experiment uses a within-subject design.

Method. For the Reversed Animals experiment, eight subjects were run in the standard Hard Animals experiment (to establish continuity with the previous experiment for this subject group) which used the inward projecting mask. Sixteen subjects were run in the same task except that the outward projecting mask was used in place of the inward projecting one. For the Within Animals experiment, the straight and curved line noise masks were replaced with a single mask which combined half straight and half curved noise. Subjects then saw the all of the animal patterns once with the inward projecting mask and once with the outward projecting mask.

Results. The results of the Reversed Animals experiment were that the subjects who searched for animals in outward projecting parts noise identified about twice as many animals as the original Hard Animals subjects (0.24 vs 0.10), but about the same as the comparison group given the Hard Animals task (0.23). Neither the inward nor outward projecting groups improved over blocks. This suggested that whatever impasses we were observing before were motivational and not cognitive.

The results of the Within Animals experiment were that subjects responded faster to the outward projecting problems than to the inward projecting ones (57 seconds vs 38 seconds), but the accuracy on the two types of problems was the same (0.32 vs 0.38, respectively) and greater than chance.

LCD Experiment: September 1987

The LCD experiment looked at transfer of learning in a diagnostic reasoning task. The subjects were to diagnose a "fault" in a display resembling an LCD numeral display. In each problem in this series, a

simulated fault caused one or more segments of the seven-segment display either to be always on, always off, or reversed: off when it should be on and on when it should be off. The subjects, by calling for the display of digits from 0 to 9, were to determine which segment(s) were affected and by which fault. Two transfer conditions and one control condition were used to determine whether learning on a more simple version of the task would produce negative transfer to a more complex version.

Method. Fifteen subjects were divided into three conditions. All subjects participated in two experimental sessions. In the first condition, subjects performed a simple version of the task on the first session and then transferred to the full task on the second session. The simple version used problems which had only one affected segment, which was either always on or always off. In the full version of the task, problems could have either one or two affected segments and could be reversed, always on, or always off. In the second condition, subjects performed a task which was more complex than the simple task, but less complex than the full task, before transferring to the full task. In this moderately complex task, problems had only one affected segment, but it could be always on, always off, or reversed. On their second session, these subjects performed the full task. Finally, the third condition received the full task on both sessions. The dependent variable was the proportion of correct responses (both segment and disease correct).

Results. Difference scores between proportion correct on first and second sessions were calculated for each subject. The mean values were -.108 for the first condition, -0.010 for the second condition, and 0.030 for the third condition. Bonferroni *t*-tests revealed that subjects who experienced the simple version of the task in the first session showed significant negative transfer relative to those who experienced the full task ($p < .05$) but that those experiencing the moderately complex task in the first session did not show significantly more negative transfer ($p > .05$).

Spectrogram Learning Experiments: November 1987 - June 1989

We shared with the ONR technical monitor the belief that the LCD studies were not as interesting a direction to pursue as the more perceptual possibilities we were considering and therefore ceased experimentation in this

line. The remainder of our studies used artificially produced voice spectrograms, displays in which time was plotted on the x axis and frequency on the y axis, with darkness of a position showing the amount of sound energy of that frequency present at that time. Figure 4 shows an example of the type of display that we used.

Nine experiments were run using pseudo-speech spectrograms as stimuli. The first studies used a scaling methodology to try to determine which visual dimensions of vowel patterns naive subjects would attend to (Vowel Scaling experiment and Scale-Learn-Scale experiment). This was followed by experiments which looked at the learning of vowel patterns

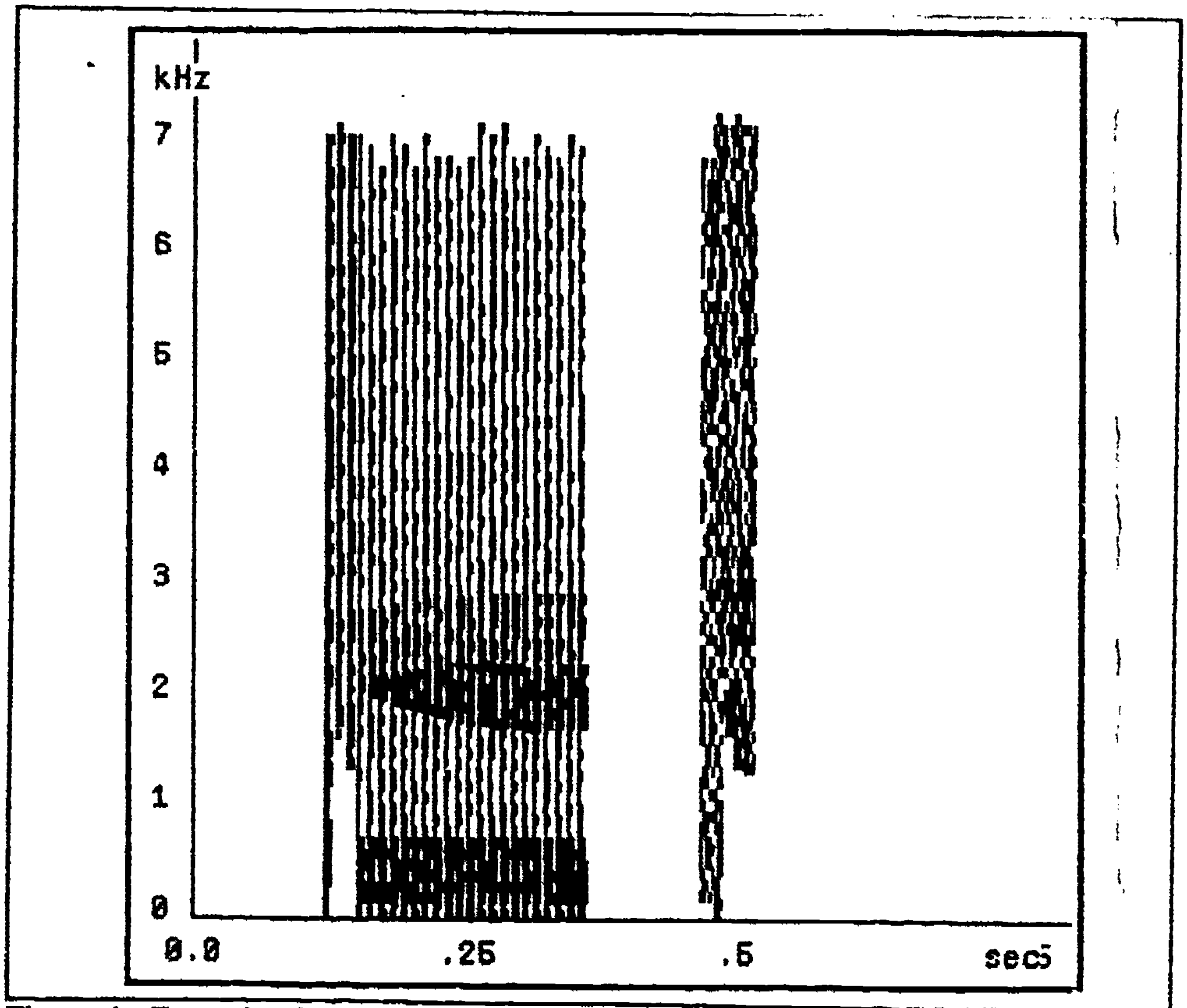

Figure 4. Example of artificial speech spectrogram.

(Vowel Transfer experiment), real word patterns (Real Word learning experiment), and finally consonant patterns (Consonant Discrimination experiments I, II, and III). A small experiment was also performed which tried to examine the influence of subjects' conceptual understanding of speech on their spectrogram reading performance (Instructional Model experiment).

To understand the logic of the experiments, a few facts about speech spectrograms are worth noting. There are two types of phonemes, vowels and consonants. Vowels consist primarily of sound energy clustered into three main frequency bands, and these bands stay at about the same frequency for a relatively long time. Consonants, on the other hand, tend to involve faster changes in frequency and somewhat less clustering around a small number of core frequencies, called *formants*. This substantial difference in appearance makes it highly likely that even a naive viewer will parse a spectrogram display into regions demarcated by phoneme boundaries. Critically important to our design is the fact that some consonants are indistinguishable from one another if one looks only at the part of the spectrogram associated with the temporal duration of the consonant. Rather, these consonants must be distinguished by examining the effects of the lip and mouth movements they involve on either preceding or following vowels. In particular, /d/ and /g/ are distinguished by their effects on the vowel which follows them, either "pulling" the start of the second and third formants together to the point of overlap or not.

This has two effects. First, vowel displays vary depending on the consonant context in which they appear. However, there are certain aspects to vowel displays that are constant. These become the critical features for identifying vowels. For identifying consonants, on the other hand, one must consider not only the part of the display showing the consonant's acoustic effect but also the neighboring vowel. Further, what is noise with respect to vowel identification is critical to neighboring consonant identification. So, identifying certain consonants like /d/ and /g/ requires noticing that part of the neighboring vowel context is relevant and, in particular, that the relevant part is the part that is more or less irrelevant to vowel identification.

We expected that impasses would occur whenever perceptual learning tasks involved distinguishing syllables that differed in whether they began

with /d/ or /g/, because the needed information for deciding on the distinction was spread over two different regions of the display and because the vowel context information needed was the "noise" with respect to vowel identification. The series of studies we conducted included some in which we tried to gather baseline data on feature salience and others in which we looked directly for the impasse effect.

Vowel Scaling Experiment and Scale-Learn-Scale Experiment

The scaling experiments were, in essence, baseline studies. A computer program was written to generate pseudo-speech spectrogram patterns based on feature descriptions of real spectrograms. The first patterns generated were vowels in a standard form (no distorting consonant context, horizontal formants) and in a transformed form (curved formants as would result from consonants immediately before or after). To compare how similarly subjects would regard the transformed and the standard vowel formants, two scaling studies were done. In the first, subjects saw all pairwise combinations of 11 vowels in standard and transformed form and rated the similarity of each pair on a numerical scale. These values were entered into a multidimensional scaling analysis. In the second experiment, a different group of subjects made similarity judgments on the 11 standard vowel patterns, then learned to distinguish the patterns, and finally, scaled the patterns again. This was done to see whether learning would change how subjects saw the patterns.

Method. In the first scaling experiment subjects scaled all pairwise combinations of 22 patterns (11 standard and 11 transformed for a total of 231 pairs). Each pair appeared on a computer screen along with a scale ranging from 1 (not similar) to 7 (very similar). Nineteen subjects rated the similarity of the 231 pairs.

In the second experiment, five subjects rated the similarity of 55 pairs of vowels (pairwise combinations of the 11 standard vowels), then learned to identify the different vowels, and finally rated them again. The rating procedure was the same as in the Vowel Scaling experiment. The learning procedure had subjects view the 11 vowels in a random order and select the name of the vowel from a screen menu. If the response was incorrect, the

subject was given the correct name. The measure of learning was the number of times the subject had to go through the list before getting them all right.

Results. The data were scaled using ALSCAL, a nonmetric, multidimensional scaling program, and INDSCAL, a related program that also examines differences between individual subjects' data. For the simple scaling experiment, the most meaningful ALSCAL solution was found with three dimensions. However, the stress value of this solution was 0.267 indicating that it was not a very good fit. Nevertheless, this solution tended to separate the patterns according to whether they were standard or transformed, whether they were low or high vowels (second formant height), and whether the formants were transformed by a slight bending (such as that which occurs when a vowel follows a bilabial stop) or by a convergence of the second and third formants (such as that which occurs when a vowel follows a velar stop).

For the Scale-Learn-Scale experiment, the scaling of the first rating achieved a stress of 0.199 in three dimensions, but only two of those dimensions, second formant height and vowel width, were readily interpretable. An INDSCAL solution indicated that most of the subjects weighted second formant height higher than both vowel width and the uninterpreted third dimension. On the Learning task, subjects took an average of 16.4 attempts to learn the 11 vowels. After learning, the subjects again rated the similarity of the vowels. On this second rating, their scaling solution looked similar to the first one. The three dimensional solution achieved a stress of 0.184 and again the recognizable dimensions were second formant height and vowel width. An INDSCAL solution was found for this second scaling and a comparison of the two revealed that most subjects increased their weighting of second formant height and decreased their weighting of vowel width. This indicates that learning may have sensitized them to using the second formant as a basis for discrimination and thus caused them to become less sensitive to the information that might help in distinguishing a prior consonant like /d/ or /g/.

Vowel Transfer Experiment

One way people might be taught to recognize vowel patterns is by training them on the standard vowel forms (which are never encountered when "reading" spectrograms of continuous speech) and expecting this training to transfer to the transformed cases the learner will encounter. It is also reasonable to expect this might not work. If subjects attend to the wrong aspects of the standard form, or don't recognize the transformed vowel as an exemplar of the standard form, no transfer would be expected. The Vowel Transfer experiment was designed to see whether this expectation was reasonable. The experiment compared transfer from the standard vowel patterns to the transformed vowel patterns with transfer in the opposite direction.

Method. Eight subjects were divided into two groups of four. One group was given the task of learning the standard vowels followed by the task of learning the transformed vowels. The second group received the same tasks but in the reverse order. The learning tasks were the same as the one described in the Scale-Learn-Scale experiment. Subjects saw 11 vowels one at a time in random order and learned to identify them by selecting their names from a screen menu. If subjects were wrong, they were told which answer was correct. The learning criterion was one errorless pass through the 11 vowels.

Results. Subjects in the first condition, who learned the standard vowels first, took an average of 28 blocks to learn the first set of vowels and an average of 7.25 blocks to learn the second. Subjects in the second condition, who learned the transformed vowels first, took an average of 11.25 blocks to learn the first task, and also took an average of 11.25 blocks to learn the second task. Learning to discriminate the transformed vowels was easier than learning to discriminate the standard vowels, likely because the transformed vowels are less similar to each other. However, learning the transformed vowels first produced a savings of 16.75 blocks on learning the standard vowels, while learning the standard vowels first only produced a savings of 4 blocks on learning the transformed vowels.

Real Word Learning Experiment

The Real Word Learning experiment examined the learning of English words made up of a stop consonant followed by a vowel followed by another stop consonant. A pseudo-spectrogram pattern was displayed on the screen and subjects were free to type in any word they chose as a response. The computer was programmed to detect alternate spellings of the target word and provided feedback when subjects made an error.

Method. Nine subjects were shown as many words as time permitted in a two hour experiment session (at least 110 and as many as 160). One subject's data was excluded because he was not a native English speaker. The subjects were free to respond with whatever word they wished, but most of them quickly learned the three letter nature of the patterns. The subjects' performance was examined by looking at the total number of correct phonemes in intervals of 10 trials.

Results. The general result was that the subjects showed quick initial learning which appeared to level off at less than perfect performance. Assuming subjects quickly learned the set of possible responses from the feedback they were given (i.e., that there were only six possible consonants and six possible vowels), two subjects showed chance performance with no improvement. The remaining six subjects each showed either abrupt or gradual initial improvement which reached a plateau between 50% and 75% correct. Looking at how subjects performed on individual phonemes revealed that /b/ and postvocalic /p/ were learned fairly quickly, followed by /d/, /t/, and prevocalic /p/, but most subjects had difficulty learning to identify /k/ and /g/. What these two patterns had in common was that they were identical to another letter (/k/ was identical to /t/ and /g/ was identical to /d/) except for their effect on the adjacent vowel. Most stops cause the formants of an adjacent vowel to curve slightly down at the consonant-vowel boundary, but the velar stops /k/ and /g/ cause the second and third formants of the vowel to curve together and meet at the consonant-vowel boundary. Subjects apparently had difficulty establishing that this difference could signal the distinction between /d/ and /g/ or /t/ and /k/.

To establish that full learning would eventually occur on this task (i.e., that subjects were not at a permanent impasse), an additional subject was

run for a total of seven consecutive sessions (1113 trials) and showed steady initial improvement for the first two sessions which appeared to level off during the third and fourth sessions before resuming to ceiling performance. This finding suggests that although learning appeared to plateau early for the first group of subjects, it would likely resume improving until it reached ceiling. This plateau appears to be due to the difficulty distinguishing the /d/ patterns from the /g/ patterns and the /t/ patterns from the /k/ patterns. This finding inspired the Consonant Discrimination Learning experiments which are described below.

Instructional Model Experiment

The purpose of this pilot experiment was to see if we could improve subjects' ability to learn to read the real word spectrograms by giving them information about how speech sounds are made and what components of the speech signal are represented in the spectrogram pattern. We looked at two types of knowledge: conceptual knowledge about how speech sounds are made, and specific cue knowledge about which spectrogram features are important for discriminating certain sounds.

Method. Thirty-two subjects were divided into four groups. These groups were: Cue Alone, Model Alone, Separate Model and Cue, and Integrated Model and Cue. The groups differed according to the verbal instructions given to the subjects. In the Cue Alone condition, subjects were shown a table which distinguished the six stop consonants and six vowels by visual features of their spectral representation. These cues included striation (voicing), width (duration), dark spots (formants), dark band height (place of articulation), and dark band curving (coarticulation effects). The subjects were told how they could use these cues to distinguish the consonants and vowels. In the Model Alone condition, subjects were shown a table which distinguished the consonants and vowels according to articulatory features (listed in parentheses above), but verbal instructions did not relate these features to any visual spectrogram features. In the Separate Model and Cue condition, subjects received all of the information in the Model Alone and Cue Alone Conditions, but this information was not related together in the verbal instructions. Finally, in the Integrated Model and Cue condition, all of the model and cue information was given and tied together in the verbal

instructions. After receiving these instructions, subjects were given the Real Word Learning experiment previously described. Subjects viewed a total of 74 words. Their performance on the first 10 words and the last 10 words was measured. On the intervening problems, subjects had access to a help window which displayed the tables they had seen during instruction. The difference between their performance on the first 10 trials and the last 10 trials was used as a measure of their improvement.

Results. The mean number of phonemes correctly identified on the first 10 problems over all subjects was 4.53. Because subjects knew that there were only six possible responses for each of the three phonemes in a pattern, chance performance on a block of 10 trials was 5.0 phonemes. A t-test showed that this first block performance was not better than chance $t(31)=1.49$, $p > .05$; and none of the means for the four instructional conditions deviated significantly from the others (range was 4.12 to 5.0). The mean number of phonemes correctly identified on the last 10 problems over all subjects was 11.16. An analysis of variance was performed to compare whether the difference in first and last block performance varied with condition. The analysis found that although significant learning occurred between the first and last block, $F(1,24)=40.96$, $p < .001$, this improvement was equal for all instructional conditions, $F(3,24)=0.98$, $p > 0.40$.

One other measure of interest was the number of times subjects in each condition used the help screen. The results showed that subjects in the Model Alone condition used the help screen the least, an average of 4.75 times. Subjects in the Cue Alone and Integrated Model and Cue condition used the facility the same amount, an average of 8.78 and 8.75 times respectively. The subjects in the Separate Model and Cue condition used the help facility the most, an average of 10.38 times. These values may reflect how useful the subjects in these conditions thought the help information was, but this did not appear to affect their learning very much.

The conclusion of this study was that no instructional effect was found for this task. The reasons are not clear, but it is likely that subjects did not adequately learn the instructional material and could not make use of it during practice. No effort was made to assess the extent of their learning of the instructional material, so this explanation is unverified.

Consonant Discrimination Learning Experiment I

In the Real Word Learning experiment, it was observed that subjects had more difficulty learning consonants which had to be distinguished by a vowel feature (formant curvature). The first Consonant Discrimination Learning experiment was undertaken to test whether this was a real effect, or whether it was due to the unequal number of consonants in each of the learning blocks. The basic design of this experiment was the same as the Real Word Learning experiment; but subjects were given all C-V-C combinations of the consonants and vowels, and they were not told of any relationship between patterns and real words. Subjects responded by selecting consonant and vowel names from a menu rather than typing in the word. Feedback was provided on error trials.

Method. Ten subjects were shown pseudo-spectrogram patterns of all CVC combinations of the consonants /b/, /p/, /d/, /g/, /t/, /k/ and vowels /i/, /e/, /ae/, /ɔ/, /u/, /o/. This produced 216 patterns, which were shown over three to four sessions. The patterns were divided into blocks of twelve, so that each consonant appeared in prevocalic and postvocalic form twice, and each vowel appeared twice. The presentation of these blocks and the order of patterns within a block was randomized. Subjects were also questioned verbally about their hypotheses and intuitions about the task. The stimuli were drawn so that /b/ and /p/ appeared similar but could be distinguished by more than one feature (such as texture and shading); /t/ and /k/ appeared similar but could be distinguished by a single feature (number of dark spots inside their pattern); and /d/ and /g/ appeared identical but could be distinguished by the curving of the adjacent vowel's formants (/g/ caused the formants to curve together). The block on which subjects learned to distinguish each of these three pairs was the main dependent variable.

Results. Subjects were considered to have learned a pair if they responded correctly on four consecutive blocks with only one error. Of the 10 subjects, 9 learned the /b/-/p/ distinction, 6 learned the /t/-/k/ distinction, and 2 learned the /d/-/g/ distinction. McNemar's exact test for correlated proportions showed that significantly more people learned the /b/-/p/ distinction than learned the /d/-/g/ distinction ($p < .02$), but the test of

whether more people learned the /t/-/k/ distinction than learned the /d/-/g/ distinction was non-significant ($p=.10$). A matched pairs sign test was used to test which distinctions were learned earlier than the others. This test revealed that the /b/-/p/ and /t/-/k/ distinctions were learned earlier than the /d/-/g/ distinction ($p < .01$ and $p < .02$ respectively).

These results appear to have verified the previous findings. It was more difficult to learn a discrimination if the critical feature is in another part (in a vowel in this case). However, it is not certain whether this effect is due to segmentation, the salience of the cues, or some other factor. The third Consonant Discrimination Learning experiment followed up this question.

Consonant Discrimination Learning Experiment II

The next Consonant Discrimination Learning experiment looked at whether the random noise added to the spectrogram patterns had any influence on the difficulty of learning the patterns. Presumably, if people are biased towards looking within a part for a feature which will identify it, then the presence of random noise will supply more hypotheses for them to consider than if the random noise were not present. The task in this experiment was simplified by using only the /d/-/g/ and /t/-/k/ consonant distinctions and only one consonant in each pattern. The presence of noise (random edging) was varied between subjects.

Method. The patterns shown to subjects were all C-V combinations of the consonants /d/, /g/, /t/, /k/ and the vowels /i/, /o/, /ae/, /e/. The 16 different patterns were shown 18 times for a total of 288 trials. In the no-noise condition, these patterns appeared with straight edges. In the noise condition, the lengths of the lines used to draw the pattern were set to a random number within about 6 mm from a set ending point. For both conditions, the problems were divided into blocks of four, where each consonant and vowel appeared once. The subjects responded separately to the consonant and vowel by selecting the symbol for each from a screen menu. The major dependent variable was the block on which a subject learned the /d/-/g/ and /t/-/k/ distinctions. Twelve subjects were run to obtain 4 full or partial learners in each condition.

Results. All four non-learners were in the noise condition. In the no-noise condition, three of the subjects learned the /t/-/k/ distinction before the /d/-/g/ distinction. In the noise condition, two subjects learned the /t/-/k/ distinction first, and two learned the /d/-/g/ distinction first. Not enough subjects were run to perform any statistical tests. The results do appear to suggest that the addition of the random noise made the task somewhat more difficult to learn.

Consonant Discrimination Learning Experiment (Selection Task)

Another question that occurred to us was whether the subjects learned the /d/-/g/ distinction last simply because it was more difficult, or whether they had to learn all of the other distinctions first to eliminate other features from consideration. Would we still find this same learning order if subjects could select which stimulus patterns they could see? To test this, we set up an experiment in which a subject responded to one block of trials in the same way as in the previous experiment, but then for the next block of trials could select which patterns to see by selecting the appropriate phonemes.

Method. It was necessary to run only one subject on this mixed presentation/selection task.

Results. The basic result is that the subject learned the /b/-/p/ distinction first, but then focused on the /d/-/g/ distinction and learned it before the /t/-/k/ distinction.

Consonant Discrimination Learning Experiment III

The final Consonant Discrimination Learning experiment tried to discover whether the learning difficulty associated with the vowel transformation cue was attributable to segmentation or some other factor such as salience. This experiment used a complex design to control for salience and task demands, but used the same task as the Noise condition in the second Consonant Discrimination experiment.

Method. To control for any differences in cue salience, each type of cue, the formant curving cue (/d/-/g/ distinction) and the number of

formants cue (/t/-/k/ distinction) was presented both within the phoneme being learned and outside of it in another part. Because this could not be done using a within subjects design, an incomplete blocks design was used. A pair of subjects provided one observation for both cues presented within and outside of a part. Thus, any difference in salience between the two cues should equally affect within and between object discriminations. To control for any task demands which may be produced by associating different parts of the pattern with different responses, subjects made a single consonant response to the whole pattern and never made a separate response for vowels. However, half of the subject pairs were given instructions biasing them to look at either the consonant or vowel (whichever contained the within object cue). Trials were divided into 8 problem blocks with each consonant represented twice and each of four vowels represented once. The block on which a subject learned one of the consonant distinctions was the major dependent variable.

Results. Subjects were considered to have learned a consonant distinction if they were correct on two consecutive blocks with one allowed error on the second block. Eighteen of the subjects learned both the within part distinction and the between part distinction, 13 learned only the within part distinction, 5 learned only the between part distinction, and 12 learned neither distinction and were not included in the analysis. Matched pairs sign tests were performed to determine which distinctions were more difficult. These tests revealed that the number of formants cue was more difficult to learn than the formant curving cue when the cues were between parts, but there was no difference between the two cues when they were within a part. This indicates that segmentation interacts with cue salience to produce learning difficulty.

However, the pattern of these results did not reproduce those reported in the first Consonant Discrimination Learning experiment. This is most likely due to the change in the task. Subjects in the previous experiment responded to both consonants and vowels, but subjects in the present experiment only made a consonant response to the whole pattern. Subjects making the vowel response likely thought the formant curving was relevant to vowel identity and failed to use it to distinguish the consonants. When the necessity of making a vowel identification was removed, subjects could consider any feature relevant to the consonant identity.

The results of this experiment indicate that subjects may be biased towards searching within a part for its distinguishing features and that this bias may be enhanced when other task demands make use of any between part cues.

Conclusions

The studies performed, and other pilot efforts with similar outcomes, make it clear that a significantly different approach will be needed if progress is to be made on impasses in perceptual learning. We did try other approaches, including extensive taking of protocols and probing for hypotheses al out what characterized various displays. However, we were not able to gain sufficient control over the generation of impasses to have them occur reliably, for most of our subjects, and over multiple experiments. Yet there were, along the way, striking examples of extended periods in which little or no learning took place.

For example, in some of our studies that showed impasses, at least temporarily, we were able to fit individual subjects' data with models that claimed performance to be constant at one level until it rose, rather quickly, to a second level. This type of model is relatively consistent with the Zeaman and House (1963) representation of learning as consisting of a period in which there is a search for relevant features followed by rapid learning of the mappings of those features onto categories. Figure 5 shows the data for one student on Consonant Discrimination Learning Experiment I. The problem was not that we never got such nice impasse patterns; rather it was that we never gained control over when they would appear. Indeed, the same experiment yielded protocols supporting the difficulty subjects had in noticing feature clusters that crossed meaningful unit (phoneme) boundaries.

We conclude that the best available tools for studying impasses in learning are probably th ls used in comparative expertise ("expert-novice") research, rather than those of the learning study. That is, one must find natural situations in which impasses occur over periods of extended learning practice and carefully assess performance at benchmark points in the course of such apprenticeship. Independent of circumstances, the time one can

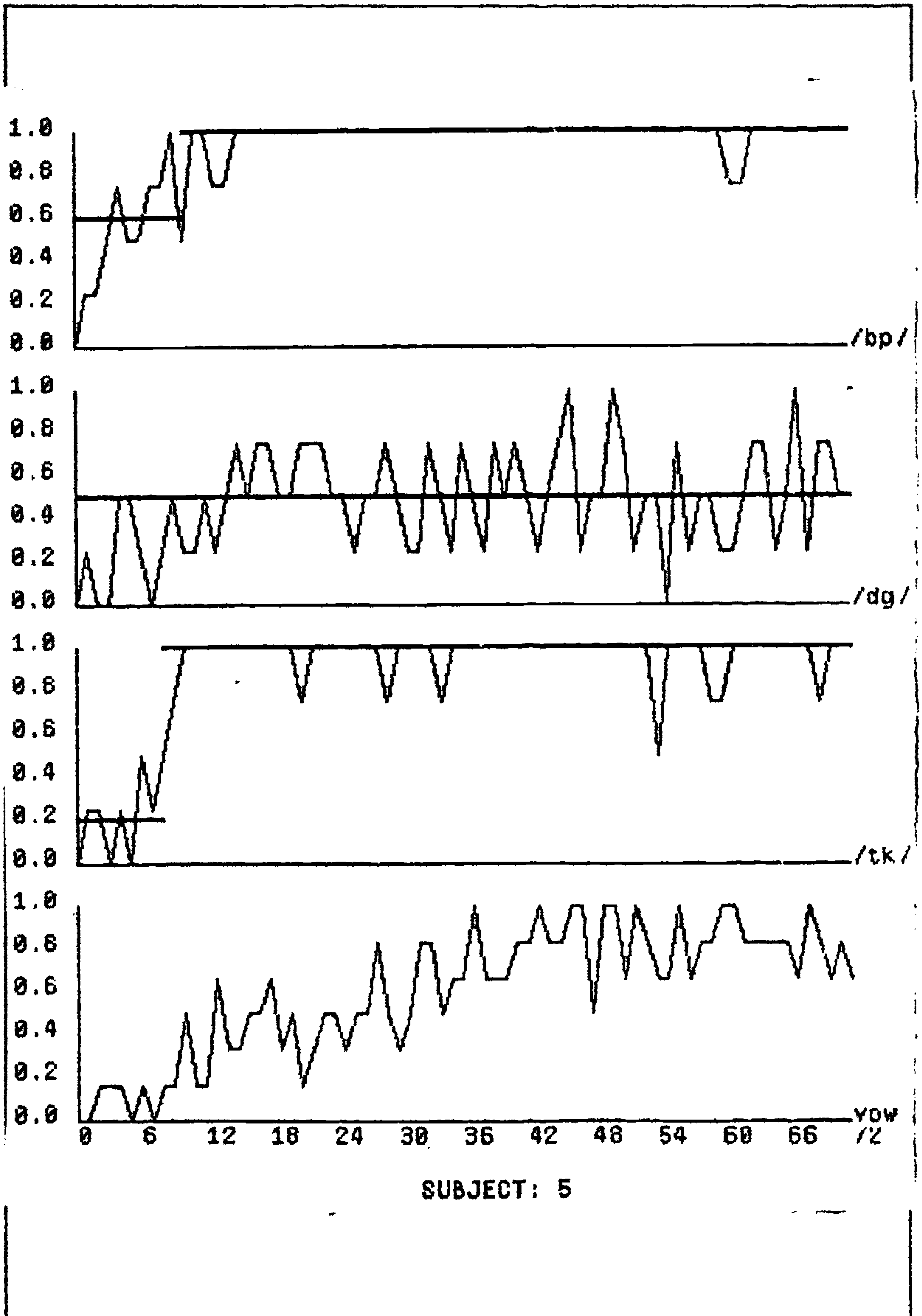

Figure 5. Example of model fits for one subject.

have to work with a research subject is always limited, and for the present purpose, it should be invested in understanding a current state of knowledge rather than trying to induce a new state that may take too long to appear. In a sense, then, the original radiological expertise studies may have been closer to the right approach than the work undertaken in the present project.

We did demonstrate impasses, though, and our views of why they occur and how they might be overcome still seem reasonable. Specifically, impasses arise when the relevant features of a situation are not apparent. Because feature noticing is extremely well developed in humans, this problem generally arises only when (a) the features defining a category are tied, by the Gestalt rules and prior knowledge of the environment, more closely to features relevant to other domain tasks than to each other; (b) a mental model of how the displays come to look they way they do has not been acquired or is not mentally manipulable with facility; and (c) no advice (rules) on how to parse the display have been acquired. Some of the displays that arise in modern technological application have these characteristics. Further, because the display forms are designed by experts, no one may notice that they have the shortcomings just mentioned.

Available Software and Data

Longer reports of each of the experiments described above, including photocopies of the display screens, are available without charge to any researcher on the ONR cognitive science mailing list. Other researchers will be accommodated but may have to pay reproduction costs if supplies run out. Similarly, the Interlisp software to produce the stimuli and run the experiments is also available under the same terms. A technical report describing the last few studies is being issued simultaneously with this final report. Address all inquiries to Alan Lesgold, LRDC, University of Pittsburgh, Pittsburgh, PA 15260.

References

Billman, D., & Heit, E. (1988). Observational learning from internal feedback. *Cognitive Science, 12*, 587-625.

Bowerman, M. (1982). Starting to talk worse: Clues to language acquisition from children's late speech errors. In S. Strauss (Ed.), *U-shaped behavioral growth*, pp. 37-61. New York: Academic Press.

Fisher, M. A., & Zeaman, D. (1973). An attention-retention theory of retardate discrimination learning. In N. R. Ellis (Ed.), *International Review of Research in Mental Retardation, Vol. 6.* New York: Academic Press.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts.

Karmiloff-Smith, A. (1979). Micro- and macrodevelopmental changes in language acquisition and other representational systems. *Cognitive Science, 3*, 91-118.

Karmiloff-Smith, A. & Inhelder, B. (1974/1975). If you want to get a head, get a theory. *Cognition, 3*, 195-212.

Keller, F. (1958). The phantom plateau. *Journal of Experimental Analysis of Behavior, 1*, 1-14.

Klahr, D. (1982). Nonmonotone assessment of monotone development: An information processing analysis. In S. Strauss (Ed.), *U-shaped behavioral growth.* New York: Academic Press.

Lesgold, A.M. (1984a). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory: Essays in honor of Gordon Bower.* San Francisco, W. H. Freeman.

Lesgold, A.M. (1984b). Human skill in a computerized society: Complex skills and their acquisition. *Behavior Research Methods, Instruments & Computers, 16*, 79-87. [Presidential address to the society for Computers in Psychology.]

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In Chi, M. T. H., Glaser, R., and Farr, M. (Eds.), *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276-298.

Richards, D.D., & Siegler, R. S. (1982). U-shaped behavioral curves: It's not whether you're right or wrong, it's why. In S. Strauss (Ed.), *U-shaped behavioral growth*, (pp. 37-61). New York: Academic Press.

Stavy, R., Strauss, S., Orpaz, N., & Carmi, G. (1982). U-shaped behavioral growth in ratio comparisons. In S. Strauss (Ed.), *U-shaped behavioral growth.* New York: Academic Press.

Strauss, S. & Stavy, R. (1982). U-shaped behavioral growth: Implications for theories of development. In W. W. Hartup (Ed.), *Review of child development research (vol. 6)*, pp. 547-599. Chicago: University of Chicago.

Zeaman, D., & House, B. J. (1963). The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill.

Dr. Edith Ackermann
Media Laboratory
E15-311
20 Ames Street
Cambridge, MA 02139

Dr. John R. Anderson
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. Stephen J. Andriole, Chairman
Department of Information Systems
and Systems Engineering
George Mason University
4400 University Drive
Fairfax, VA 22030

Prof. J Annett
University of Warwick
Department of Psychology
Coventry CV4 7AL
ENGLAND

Dr. Gary Aston-Jones
Department of Biology
New York University
Rm. 1009 Main Bldg.
100 Washington Square East
New York, NY 10003

Dr. Patricia Baggett
School of Education
610 E. University, Rm 1302D
University of Michigan
Ann Arbor, MI 48109-1259

Dr. Gautam Biswas
Department of Computer Science
Box 1688, Station B
Vanderbilt University
Nashville, TN 37235

Dr. Lyle E. Bourne, Jr.
Department of Psychology
Box 345
University of Colorado
Boulder, CO 80309

Dr. Robert Calfee
School of Education
Stanford University
Stanford, CA 94305

Dr. Joseph C. Campione
Center for the Study of Reading
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Jaime G. Carbonell
Computer Science Department
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. John M. Carroll
IBM Watson Research Center
User Interface Institute
P.O. Box 704
Yorktown Heights, NY 10598

Dr. David E. Clement
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Jere Confrey
Cornell University
Dept. of Education
Room 490 Roberts
Ithaca, NY 14853

Dr. Lynn A. Cooper
Department of Psychology
Columbia University
New York, NY 10027

Dr. Thomas E. DeZern
Project Engineer, AI
General Dynamics
PO Box 748/Mail Zone 2646
Fort Worth, TX 76101

Dr. Ronna Dillon
Department of Guidance and
Educational Psychology
Southern Illinois University
Carbondale, IL 62901

Dr. J. Stuart Donn
Faculty of Education
University of British Columbia
2125 Main Mall
Vancouver, BC CANADA V6T 1Z5

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
(12 Copies)

Dr. John Ellis
Navy Personnel R&D Center
Code 51
San Diego, CA 92252

ERIC Facility-Acquisitions
2440 Research Blvd, Suite 550
Rockville, MD 20850-3238

Dr. K. Anders Ericsson
University of Colorado
Department of Psychology
Campus Box 345
Boulder, CO 80309-0345

Dr. Marshall J. Farr, Consultant
Cognitive & Instructional Sciences
2520 North Vernon Street
Arlington, VA 22207

CAPT J. Finelli
Commandant (G-PTE)
U.S. Coast Guard
2100 Second St., S.W.
Washington, DC 20593

Dr. Norman Frederiksen
Educational Testing Service
(05-R)
Princeton, NJ 08541

Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2E9

Dr. Robert M. Gagne
1456 Mitchell Avenue
Tallahassee, FL 32303

Psychology Department
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Marvin D. Glock
101 Homestead Terrace
Ithaca, NY 14856

Dr. Dwight J. Goehring
ARI Field Unit
P.O. Box 5787
Presidio of Monterey, CA 93944-5011

Dr. Sherrie Gott
AFHRL/MOMJ
Brooks AFB, TX 78235-5601

Dr. Wayne Gray
Artificial Intelligence Laboratory
NYNEX
500 Westchester Avenue
White Plains, NY 10604

H. William Greenup
Dep Asst C/S, Instructional
Management (E03A)
Education Center, MCCDC
Quantico, VA 22134-5050

Dr. Gerhard Grossing
Atominstitut
Schuttelstrasse 115
Vienna
AUSTRIA A-1020

Stevan Harnad
Editor, The Behavioral and
Brain Sciences
20 Nassau Street, Suite 240
Princeton, NJ 08542

Dr. James Hiebert
Department cf Educational
Development
University of Delaware
Newark, DE 19716

Dr. Steven A. Hillyard
Department of Neurosciences, M-008
University of California, San Diego
La Jolla, CA 92093

Ms. Julia S. Hough
Cambridge University Press
40 West 20th Street
New York, NY 10011

Dr. William Howell
Chief Scientist
AFHRL/CA
Brooks AFB, TX 78235-5601

Z. Jacobson
Bureau of Management Consulting
701-365 Laurier Ave., W.
Ottawa, Ontario K1H 5W3
CANADA

Dr. Daniel B. Jones
U.S. Nuclear Regulatory
Commission
NRR/ILRB
Washington, DC 20555

Mr. Paul L. Jones
Research Division
Chief of Naval Technical Training
Building East-1
Naval Air Station Memphis
Millington, TN 38054-5056

Dr. Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

Dr. Daniel Kahneman
Department of Psychology
University of California
Berkeley, CA 94720

Dr. Ruth Kanfer
University of Minnesota
Department of Psychology
Elliott Hall
75 E. River Road
Minneapolis, MN 55455

Dr. Michael Kaplan
Office of Basic Research
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Milton S. Katz
European Science Coordination
Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1500

Dr. Steven W. Keele
Department of Psychology
University of Oregon
Eugene, OR 97403

Dr. Frank Keil
Department of Psychology
228 Uris Hall
Cornell University
Ithaca, NY 14850

Dr. Wendy Kellogg
IBM T. J. Watson Research Ctr.
P.O. Box 704
Yorktown Heights, NY 10598

Dr. Thomas Killion
AFHRL/OT
Williams AFB, AZ 85240-6457

Dr. J. Peter Kincaid
Army Research Institute
Orlando Field Unit
c/o PM TRADE-E
Orlando, FL 32813

Mr. David A. Kobus
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92138

Dr. Sylvan Kornblum
University of Michigan
Mental Health Research Institute
205 Washtenaw Place
Ann Arbor, MI 48109

Dr. Gary Kress
628 Spazier Avenue
Pacific Grove, CA 93950

Dr. Lois-Ann Kuntz
3010 S.W. 23rd Terrace
Apt. No. 105
Gainesville, FL 32608

Dr. Pat Langley
NASA Ames Research Ctr.
Moffett Field, CA 94035

Dr. Robert W. Lawler
Matthews 118
Purdue University
West Lafayette, IN 47907

Dr. John Levine
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

Matt Lewis
Department of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. Robert Lloyd
Dept. of Geography
University of South Carolina
Columbia, SC 29208

Dr. Don Lyon
P. O. Box 44
Higley, AZ 85236

Dr. William L. Maloy
Code 04
NETPMSA
Pensacola, FL 32509-5000

Dr. Elizabeth Martin
AFHRL/OTE
Williams AFB
AZ 85240

Dr. James G. May
Department of Psychology
University of New Orleans
Lakefront
New Orleans, LA 70148

Dr. Joseph C. McLachlan
Code 52
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. D. Michie
The Turing Institute
George House
36 North Hanover Street
Glasgow G1 2AD
UNITED KINGDOM

Dr. Allen Munro
Behavioral Technology
Laboratories - USC
250 N. Harbor Dr., Suite 309
Redondo Beach, CA 90277

Prof. David Navon
Department of Psychology
University of Haifa
Haifa 31999
ISRAEL

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Librarian
Naval Center for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington, DC 20375-5000

Dr. Paul O'Rorke
Information & Computer Science
University of California, Irvine
Irvine, CA 92717

Dr. Stellan Ohlsson
Learning R & D Center
University of Pittsburgh
Pittsburgh, PA 15260

Dr. James B. Olsen
WICAT Systems
1875 South State Street
Orem, UT 84058

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Dr. Judith Orasanu
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Okchoon Park
Army Research Institute
PERI-2
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Roy Pea
Institute for Research
on Learning
2550 Hanover Street
Palo Alto, CA 94304

Dr. C. Perrino, Chair
Dept. of Psychology
Morgan State University
Cold Spring La.-Hillen Rd.
Baltimore, MD 21239

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Dr. Martha Polson
Department of Psychology
University of Colorado
Boulder, CO 80309-0345

Dr. Peter Polson
University of Colorado
Department of Psychology
Boulder, CO 80309-0345

Dr. Michael I. Posner
Department of Psychology
University of Oregon
Eugene, OR 97403

Dr. Robert B. Post
Psychology Department
University of California
Davis, CA 95616

Dr. Daniel Reisberg
Reed College
Department of Psychology
Portland, OR 97202

Dr. Ernst Z. Rothkopf
AT&T Bell Laboratories
Room 2D-456
600 Mountain Avenue
Murray Hill, NJ 07974

Dr. Arthur G. Samuel
Yale University
Department of Psychology
Box 11A, Yale Station
New Haven, CT 06520

Dr. Walter Schneider
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Hans-Willi Schroiff
Psychological Research
Henkel Corp./WSC
Henkelstr. 67
D-4000 Duesseldorf 1 FRG
WEST GERMANY

Dr. Randall Shumaker
Naval Research Laboratory
Code 5510
4555 Overlook Avenue, S.W.
Washington, DC 20375-5000

Dr. Zita M. Simutis
Chief, Technologies for Skill
Acquisition and Retention
ARI
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Saul Sternberg
University of Pennsylvania
Department of Psychology
3815 Walnut Street
Philadelphia, PA 19104-6196

Dr. Patrick Suppes
Stanford University
Institute for Mathematical
Studies in the Social Sciences
Stanford, CA 94305-4115

Dr. John Tangney
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Perry W. Thorndyke
FMC Corporation
Central Engineering Labs
1205 Coleman Avenue, Box 580
Santa Clara, CA 95052

Dr. Paul T. Twohig
Army Research Institute
ATTN: PERI-RL
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Zita E. Tyer
Department of Psychology
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. Kurt Van Lehn
Department of Psychology
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213

Dr. Shih-sung Wen
Department of Psychology
Jackson State University
1400 J. R. Lynch Street
Jackson, MS 39217

Mr. Paul T. Wohig
Army Research Institute
5001 Eisenhower Ave.
ATTN: PERI-RL
Alexandria, VA 22333-5600

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550